Journal of Biomolecular NMR, **11:** 363–380, 1998.
KLUWER/ESCOM

Multinuclear NMR resonance assignments and the secondary structure of *Escherichia coli* thioesterase/protease I: A member of a new subclass of lipolytic enzymes

Ta-hsien Lin[a,*,**], Chinpan Chen[a,**], Rong-Fong Huang[a], Ya-Lin Lee[b], Jei-Fu Shaw[b] & Tai-huang Huang[a,***]
[a]*Division of Structural Biology, Institute of Biomedical Sciences and* [b]*Institute of Botany, Academia Sinica, Nankang, Taipei, Taiwan 11529, Republic of China*

Received 21 July 1997; Accepted 3 October 1997

Key words: chemical shift indices, heteronuclear triple-resonance NMR, lipase, lipolytic enzyme, resonance assignment

Abstract

Escherichia coli thioesterase/protease I is a 183 amino acid protein with a molecular mass of 20 500. This protein belongs to a new subclass of lipolytic enzymes of the serine protease superfamily, but with a new GDSLS consensus motif, of which no structure has yet been determined. The protein forms a tetramer at pH values above 6.5 and exists as a monomer at lower pH values. Both monomer and tetramer are catalytically active. From analysis of a set of heteronuclear multidimensional NMR spectra with uniform and specific amino acid labeled protein samples, we have obtained near-complete resonance assignments of the backbone ^{1}H, ^{13}C and ^{15}N nuclei (BMRB databank accession number 4060). The secondary structure of *E. coli* thioesterase/protease I was further deduced from the consensus chemical shift indices, backbone short- and medium-range NOEs, and amide proton exchange rates. The protein was found to consist of four β-strands and seven α-helices, arranged in alternate order. The four β-strands were shown to form a parallel β-sheet. The topological arrangement of the β-strands of -1x, $+2$x, $+1$x appears to resemble that of the core region of the αβ hydrolase superfamily, typically found in common lipases and esterases. However, substantial differences, such as the number of β-strands and the location of the catalytic triad residues, make it difficult to give a definitive classification of the structure of *E. coli* thioesterase/protease I at present.

Introduction

Lipases are important enzymes in lipid metabolism and in signal transduction (Borgström and Brockman, 1984). They are unique in their ability to catalyze reactions at an oil–water interface and are valuable enzymes for the stereospecific hydrolysis and synthesis of a wide variety of technologically important regiospecific esters, organic acids, and alcohols (Santaniello et al., 1992; Cambillau et al., 1996). The lipolytic enzymes investigated so far vary considerably in size and amino acid sequence. Yet they all belong to the class of serine esterases with a 3D structure that conforms to the symbol α/β hydrolase superfamily (Ollis et al., 1992). The secondary structure of this class of enzymes is characterized by alternate α-helices and β-strands. The active site is a Ser-Asp/Glu-His catalytic triad, similar to that observed in serine proteases. The nucleophilic serine is invariably located at the center of an extremely sharp turn between a β-strand and a buried α-helix. In most lipases this serine is found in the sequence GxSxG, considered to be an identifying feature of lipases (Jaeger et al., 1994; Rubin, 1994). Recently, a new class of lipolytic enzymes with diverse substrate specificity

*Present address: Department of Medical Research, Veterans General Hospital-Taipei, Shi-Pai, Taipei, Taiwan 11217, Republic of China.
**T.-h. Lin and C.-p. Chen contributed equally to this paper.
*** To whom correspondence should be addressed.

and regiospecificity was found (Upton and Buckley, 1995). Members of this class include *Aeromonas hydrophila* lipase/acyltransferase, *Vibrio parahaemolyticus* hemolysin/phospholipase, *Xenorhabdus luminescens* lipase, *Pseudomonas putida* ORF in the *trpE-trpG* region, *Arabidopsis thaliana* proline-rich protein, *Brassica napus* proline-rich protein, *Virio mimicus* arylesterase, *E. coli* thioesterase/protease I, and *Streptomyces scabies* suberin esterase. This class of nine enzymes possesses a distinctive GDSLS sequence motif, instead of the common GxSxG motif. In *Aeromonas* sp., the enzyme was found to be inactive when the sequence was changed to GDSLG (Robertson et al., 1994). All the proteins in this group share little sequence homology with the known lipases and have at least four other sequence blocks in common which all appear in the same order (Upton and Buckley, 1995). This provides more evidence that this class of enzymes represents a new family or subfamily of lipolytic enzymes. This sequence alignment also allows the identification of a Ser-Asp-His triad as the likely active site residues, similar to that found in common lipases. The importance of the active site triad in *Vibrio mimicus* arylesterase was confirmed by a site-directed mutagenesis study (Lee et al., 1997). However, the other conserved residues in the motifs surrounding each of the presumed catalytic amino acids are not the same as those found in common lipases. This, together with the fact that the GDSLS sequence in most of the proteins in this subclass of lipolytic enzymes is much closer to the N-terminus whilst the GxSxG motif is typically near the middle of the peptide sequence in the common lipases, indicates that there may be significant differences between the structures of the two classes of lipases. At present, no structure is known for a member of this family of enzymes; thus, no information regarding their structural relationship to other lipases is available. Therefore, determining the structures of this subfamily of lipolytic enzymes has become an interesting structural biology problem.

Kass et al. (1967) first reported the presence of thioesterase activity in extracts of *E. coli*. Two distinct enzymes, a low molecular weight thioesterase I and a high molecular weight thioesterase II, that catalyze the hydrolysis of fatty acyl coenzyme A (acyl-CoA) substrates were later detected after fractionation of cell extracts (Barnes and Wakil, 1968; Barnes et al., 1970). The *tesA* gene encoding the *E. coli* thioesterase I was later cloned and sequenced (Cho and Cronan, 1993). *E. coli* protease I was first reported by Pacaud and Uriel (1971), who found that a partially purified *E. coli* protease I enzyme was able to convert the native form of polynucleotide phosphorylase into a smaller, more active form of the enzyme. This protease I enzyme was further purified and characterized (Pacaud and Uriel, 1971; Pacaud et al., 1976). The *apeA* gene encoding the protease I enzyme was later subcloned and sequenced (Ichihara et al., 1993). A comparison of the sequence of the *tesA* and *apeA* genes showed that the *E. coli* thioesterase I and protease I were the same protein.

E. coli thioesterase/protease I, a member of the new subclass of lipolytic enzymes, is a 183 amino acid protein of 20.5 kDa. It catalyzes the hydrolytic cleavage of long-chain (C_{12}–C_{18}) fatty acyl coenzyme A (CoA) thioesters *in vitro* and also cleaves fatty acyl carrier protein (ACP) thioesters at a rate of 10^3 to 10^4 times slower than acyl-CoA esters of the same length (Barnes and Wakil, 1968; Spencer et al., 1978). Thus, it terminates fatty acid chain growth during fatty acid synthesis by hydrolysis of the acyl thioester linkage between the fatty acid and phosphopantetheine cofactor of fatty acid synthesis. This protein has also been shown to possess arylesterase activity (Lee et al., 1997). Cho and Cronan (1994) further showed that the same active site is used for both fatty acid and amino acid substrates and that it behaves as a thioesterase *in vivo*. Defects in the export of this enzyme result in disrupted regulation of fatty acid synthesis (Cho and Cronan, 1995). The exact physiological role of the protein is unclear as the deletion of this gene does not seem to produce any change in growth phenotype (Cho and Cronan, 1993; Ichihara et al., 1993). The rather versatile enzymatic activities of *E. coli* thioesterase/protease I render it a good candidate for further engineering for industrial applications.

We have prepared several uniform and specific amino acid ^{13}C- and/or ^{15}N-labeled protein samples and have employed multinuclear, multidimensional NMR techniques to determine the solution structure of *E. coli* thioesterase/protease I. In this paper we report NMR resonance assignment, secondary structure identification, and the structural implication of *E. coli* thioesterase/protease I.

Materials and Methods

Protein purification

E. coli thioesterase/protease I was isolated from the BL21(DE3) strain of *E. coli* containing a plasmid which carries the *tesA*/*apeA* gene for producing thioesterase I, under the control of T7 polymerase (Lee et al., 1997; Su, 1997). A string of six histidines was added to the C-terminus to facilitate protein purification by a Ni-NTA column, as described previously (Lee et al., 1997). The histidine tag was not cleaved from the C-terminus. The purity of the protein was checked by gel electrophoresis and was found to be better than 95% after elution from the Ni column. Typical protein yields were about 15 mg/l for cells grown in M9 medium. The enzyme was checked for its thioesterase, arylesterase, and protease activities using palmitoyl-CoA, *p*-nitrophenyl esters, and *N*-carbobenzoxy-D-phenylalanine *p*-nitrophenyl ester (D-NBPNPE) and *N*-carbobenzoxy-L-phenylalanine *p*-nitrophenyl ester (L-NBPNPE), respectively, at various pH values (Lee et al., 1997). Uniform ^{13}C- and/or ^{15}N-labeled protein samples were purified from *E. coli* cells grown in M9 medium, supplemented with 2 g/l of U-^{13}C glucose and/or 1 g/l of $^{15}NH_4Cl$. To obtain samples with ^{15}N labeled at specific amino acids, *E. coli* cells were grown at 37 °C in M9 medium, supplemented with unlabeled glucose and NH_4Cl. After cells had grown to an OD of 0.8, the desired amino acid (^{15}N-labeled) was added, together with IPTG. The amount of labeled amino acid added to each preparation was 600 mg/l of alanine, 60 mg/l of leucine, and 60 mg/l of phenylalanine.

NMR spectroscopy

Samples for NMR experiments contained about 0.3 ml of 2–3 mM protein in 50 mM phosphate buffer, pH 3.5, in 5 mm Shigemi NMR tubes (Shigemi Co., Tokyo, Japan). Buffer exchange and pH adjustments were achieved by repeated dilution with appropriate buffers and concentrating with centricon filters (Millipore Co., Bedford, MA, U.S.A.). The pH of the final filtrate was taken as the pH of the protein sample. Upon attaining the desired pH, 10% (v/v) of D_2O was added. For preparing the sample in D_2O, the concentrated protein sample was repeatedly lyophilized and redissolved in D_2O. pH values were measured with a JENCO microelectronic pH-vision model 6071 pH meter equipped with a 4 mm electrode. All reported pH values were direct readings from the pH meter without correction for the isotope effect. pH* is used to indicate direct pH meter readings of D_2O solutions, uncorrected for the deuterium isotopic effect using electrodes standardized in H_2O buffers.

All experiments were performed on a Bruker AMX-600 NMR spectrometer equipped with an amplifier capable of generating shaped field-gradient pulses of 50 G/cm. A 5 mm inverse triple-resonance probe head (^{1}H/^{13}C/^{15}N), fitted with a self-shielded X,Y,Z-gradient coil, was used. Quadrature detection in the indirectly detected dimension was accomplished using the States-TPPI method (Marion et al., 1989b). Water suppression was achieved by applying the WATERGATE sequence (Piotto et al., 1992). ^{15}N or ^{13}C decoupling during acquisition was accomplished with the GARP pulse sequence (Shaka et al., 1985). Pulse sequences of triple-resonance experiments, such as HNCA, HNCO, CBCA(CO)NH, etc., were modified by applying one scrambling Z-gradient during each INEPT step, when the magnetization of interest is converted into $I_Z S_Z$ order. Gaussian-shaped ^{13}CO 90° and 180° pulses were used to select for excitation of either $^{13}C'$ or $^{13}C^{\alpha}$spins. Carbon 90° and 180° pulses applied on the same carrier frequency were adjusted to the same length using different power levels. All spectra were recorded at 310 K.

A 2D NOESY (Jeener et al., 1979) with a mixing time of 80 ms and a 2D TOCSY with a 50 ms spin-lock time were performed for unlabeled protein dissolved in D_2O. 2D ^{1}H-^{15}N-HSQC (Bodenhausen and Ruben, 1980), 3D ^{1}H-^{15}N-TOCSY-HSQC (Marion et al., 1989a) with a 55 ms isotropic mixing time, and 3D ^{1}H-^{15}N-NOESY-HSQC (Marion et al., 1989a) with 110 and 200 ms NOESY mixing periods were acquired with a ^{15}N uniformly labeled sample. A study of slowly exchanging amide protons was accomplished by acquiring a series of 2D ^{1}H-^{15}N-HSQC spectra immediately after dissolution of the lyophilized, ^{15}N uniformly labeled sample in D_2O. 2D ^{1}H-^{15}N-HSQC spectra were also acquired with samples specifically (amide-^{15}N)-labeled at the following amino acids: alanine, leucine, phenylalanine, and serine. 2D ^{1}H-^{13}C-CT-HSQC (Vuister and Bax, 1992) and 3D HCCH-TOCSY (Bax et al., 1990; Wang and Zuiderweg, 1995) spectra were acquired on a ^{13}C uniformly labeled sample in D_2O. The 3D triple-resonance experiments CT-HNCO, CT-HNCA, CT-HN(CO)CA (Grzesiek and Bax, 1992), CT-HN(CA)CO (Clubb et al., 1992), CBCA(CO)NH (Grzesiek and Bax, 1993), and HNCACB (Wittekind and Mueller, 1993) were performed on a ^{15}N,^{13}C uni-

Table 1. Acquisition parameters for experiments

Experiment	Nucleus			No. of complex points			SW (Hz)			Scans	Total time (h)
	F1	F2	F3	F1	F2	F3	F1	F2	F3		
HNCO	^{15}N	^{13}CO	$^{1}H^{N}$	32	40	512	2432	1962	4237	8	14
HNCA	^{15}N	$^{13}C^{\alpha}$	$^{1}H^{N}$	52	105	512	2432	9054	4237	16	120
HN(CA)CO	^{15}N	^{13}CO	$^{1}H^{N}$	32	40	512	2432	1962	4237	48	87
HN(CO)CA	^{15}N	$^{13}C^{\alpha}$	$^{1}H^{N}$	32	32	512	2432	5282	4237	8	12
CBCA(CO)NH	^{15}N	$^{13}C^{\alpha/\beta}$	$^{1}H^{N}$	32	64	512	2432	12072	4237	32	91
HNCACB	^{15}N	$^{13}C^{\alpha/\beta}$	$^{1}H^{N}$	32	64	512	2432	12072	4237	32	94
TOCSY-HSQC	^{1}H	^{15}N	$^{1}H^{N}$	128	64	512	7801	2432	4237	8	92
NOESY-HSQC	^{1}H	^{15}N	$^{1}H^{N}$	128	64	512	7801	2432	4237	8	133
HCCH-TOCSY	^{1}H	^{13}C	^{1}H	128	105	512	6849	10563	6849	8	141
C(CO)NH	^{15}N	^{13}C	$^{1}H^{N}$	32	70	512	2432	12072	4237	32	95
CT-HSQC	^{13}C	^{1}H		160	1024		10563	6849		48	5
HSQC	^{15}N	$^{1}H^{N}$		128	1024		2432	4237		16	0.66

formly double labeled sample in H_2O. The acquisition parameters for these experiments are listed in Table 1.

All spectra were processed using UXNMR (Bruker AG, Karlsruhe, Germany) and analyzed using AURELIA (Neidig et al., 1995) on an SGI workstation. Linear prediction was used in the ^{15}N dimension to improve the digital resolution. Chemical shifts were referenced to DSS (2,2-dimethyl-2-silapentane-5-sulfonate) at 0 ppm (Wishart and Sykes, 1994). The ^{15}N and ^{13}C chemical shifts were indirectly referenced using the following consensus Ξ ratios of the zero-point frequencies at 310 K: 0.101329118 for $^{15}N/^{1}H$ and 0.251449530 for $^{13}C/^{1}H$ (Wishart et al., 1995).

Results

The resonances of the proton 1D spectrum of thioesterase/protease I at pH 7.0 are very broad, indicating the formation of a multimer under this condition. However, the line width narrowed considerably upon lowering the pH to 6.5 and below. Gel exclusion column chromatography further confirmed that the protein forms a tetramer at pH values above 6.5 and a monomer at pH values below 6.5. An activity assay showed that both monomer and tetramer are active (Su, 1997). The spectra obtained at pH 6.5 and pH 3.5 are very similar, suggesting little conformational difference at these two pH values. In the monomeric form the protein is well behaved, with excellent temperature and pH stability. The protein also showed excellent chemical shift dispersion as indicated by the 2D ^{1}H-^{15}N HSQC spectrum of thioesterase/protease I, acquired at 310 K and pH 3.5 in 90% H_2O/10% D_2O, as shown in Figure 1. Excluding the N-terminal residue and the 13 proline residues, there are 175 potentially observable backbone NH-^{15}N cross peaks in the ^{1}H-^{15}N-HSQC spectrum of *E. coli* thioesterase/protease I (there are six additional histidine residues in the C-terminus). One hundred and sixty-nine well-resolved resonances were observed. Thirty-eight side-chain NH_2 resonances, out of 50 expected from 8 asparagine and 17 glutamine residues, were observed. In addition, all resonances of the five indole $N^{\epsilon}H$ of tryptophan residues and eight arginine side-chain $N^{\epsilon}H$ (folded) were observed.

Spin system identification

Spin system identification was accomplished by analyzing the 3D ^{15}N-TOCSY-HSQC spectrum obtained with a ^{15}N-labeled protein sample. We observed 164 out of 175 potentially observable intra-residue H^{N}-H^{α} cross peaks in the 3D TOCSY-HSQC spectrum. The 10 glycine residues were assigned easily from their characteristic high-field ^{15}N resonances and the presence of two $C^{\alpha}H$ cross peaks in the 3D ^{15}N-TOCSY-HSQC spectrum. In addition, we readily identified most spin systems for alanine, valine, and threonine residues, as well as the residues with AMX and AMPTX spin systems. Out of 25 asparagine and glutamine side-chain amides, 22 were observed in the 3D TOCSY-HSQC spectrum. The 2D ^{13}C-CT-HSQC experiment of Vuister and Bax (1992) was found to be very useful as it had good resolution in the ω_1 dimension due to the removal of ^{13}C-^{13}C coupling in the

Table 2. ^{1}H, ^{15}N and ^{13}C chemical shifts for *E. coli* thioesterase I in 90% H_2O/10% D_2O at 310 K, pH 3.5

Residue	^{15}N	NH	H^{α}	H^{β}	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	^{13}CO	Other ^{1}H	Other ^{15}N, ^{13}C
Met[1]									
Ala[2]	117.61	7.51	4.12	1.45	51.77	19.80	173.87		
Asp[3]	121.07	8.72	4.80	2.87, 2.68	54.20	41.71	174.88		
Thr[4]	117.32	8.62	5.18	3.93	62.70	71.42	173.26	$H^{\gamma 2}$ 1.05	$C^{\gamma 2}$ 21.21
Leu[5]	129.82	9.54	5.31	2.08, 1.55	53.70	44.05	173.06		
Leu[6]	128.57	8.72	5.08		53.05	46.21	174.38		
Ile[7]	126.70	9.12	4.43	2.06	60.25	37.42	172.85	$H^{\gamma 1}$ 1.11; $H^{\gamma 2}$ 0.73; $H^{\delta 1}$ 0.55	$C^{\gamma 2}$ 16.35; $C^{\delta 1}$ 12.51
Leu[8]	130.76	9.05	4.64	2.08	52.83	43.69	173.43	H^{δ} 0.61	
Gly[9]	111.70	6.87	5.52, 3.70		44.46		173.36		
Asp[10]	117.63	7.75	5.45		51.48	40.77	179.47		
Ser[11]					66.91				
Leu[12]	118.04	8.05	4.64		57.36	40.34	179.46		
Ser[13]	113.90	7.79	4.71	3.97, 3.97	60.14	64.63	172.85		
Ala[14]	119.82	8.23	3.86	1.04	53.30	19.97	177.12		
Gly[15]	102.32	7.67	3.96, 3.50		45.00		172.04		
Tyr[16]	125.14	8.26	4.37	2.57, 2.57	57.75	39.37	174.48	H^{δ} 6.73; H^{ϵ} 6.63	
Arg[17]	122.63	8.72	3.47	1.68	57.29	26.62	174.42	H^{γ} 0.23; H^{δ} 3.19; H^{ϵ} 7.26	N^{ϵ} 83.78
Met[18]	111.07	7.23	2.91		53.82	35.07	173.67		
Ser[19]	115.13	8.39	4.31	3.90, 3.81	58.00	63.62	175.49		
Ala[20]	127.30	8.77	4.03	1.44	55.28	18.21	180.47		
Ser[21]	108.88	7.86	3.95	3.82, 3.82	62.67	59.40	173.46		
Ala[22]	121.70	7.50	4.26	1.34	50.95	19.91	176.61		
Ala[23]	118.26	6.88	4.18	1.08	51.43	19.85	179.05		
Trp[24]	118.01	8.98	4.65	3.17, 3.05	61.12		174.60	$H^{\epsilon 1}$ 9.65; $H^{\delta 1}$ 7.37; $H^{\epsilon 3}$ 6.86; $H^{\zeta 3}$ 6.49; $H^{\eta 2}$ 6.76; $H^{\zeta 2}$ 7.70	$N^{\epsilon 1}$ 129.20
Pro[25]					66.07	30.96	178.13		
Ala[26]	118.88	7.20	4.00	1.46	54.32	18.27	180.17		
Leu[27]	117.32	7.13	4.10	1.67, 1.36	56.79	40.10	180.37		
Leu[28]	121.07	8.41	3.68	1.74	57.24	42.17	177.42		
Asn[29]	115.76	8.01	4.30	3.01, 2.81	57.30	39.38	177.12	$H^{\delta 2}$ 7.71, 6.86	$N^{\delta 2}$ 113.57
Asp[30]	115.76	7.72	4.44	2.84, 2.82	56.24	38.89	174.78		
Lys[31]	118.57	7.67	4.16	1.78, 1.47	58.37	33.06	178.54		
Trp[32]	117.38	8.10	4.97	3.58, 3.26	55.43	30.23	176.81	$H^{\epsilon 1}$ 10.10; $H^{\delta 1}$ 6.93; $H^{\eta 2}$ 6.71; $H^{\zeta 3}$ 6.55; $H^{\zeta 2}$ 7.24	$N^{\epsilon 1}$ 126.07
Gln[33]	118.26	8.01	4.29		58.20	28.00	175.90		
Ser[34]	114.20	8.55	4.16	3.96, 3.84	60.61	62.77	174.17		
Lys[35]	120.45	8.47	4.37	1.68, 1.31	57.63	33.92	176.51		
Thr[36]	116.07	7.85	4.60	3.64	62.11	71.83	173.06	$H^{\gamma 2}$ 1.11	$C^{\gamma 2}$ 22.31
Ser[37]	120.45	8.66	4.72	3.83, 3.83	57.93	63.74	173.46		
Val[38]	125.76	8.61	4.89	1.90	61.37	32.82	175.49	H^{γ} 0.98, 0.66	C^{γ} 20.82
Val[39]	128.57	8.88	3.96	1.42	60.68	33.49	173.97	H^{γ} 0.62, 0.12	C^{γ} 20.59, 19.06
Asn[40]	126.39	9.14	4.71	3.05, 2.14	51.39	37.92	173.97	$H^{\delta 2}$ 7.84, 6.69	$N^{\delta 2}$ 109.51
Ala[41]	132.32	8.87	4.76	1.27	50.18	20.29	176.61		
Ser[42]	116.38	7.27	4.84	3.02, 3.02	58.75	63.82	172.74		

Table 2. Continued

Residue	^{15}N	NH	H^{α}	H^{β}	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	^{13}CO	Other ^{1}H	Other ^{15}N, ^{13}C
Ile[43]	122.63	7.50	4.14	1.70	59.47	40.05	176.20	$H^{\gamma 1}$ 1.23, 0.98; $H^{\gamma 2}$ 0.80; $H^{\delta 1}$ 0.74	$C^{\gamma 2}$ 16.86; $C^{\delta 1}$ 16.35
Ser[44]	122.63	8.91	3.84		61.19	62.90			
Gly[45]	106.07	8.77	4.76, 3.86		45.19		174.88		
Asp[46]	119.51	7.27	4.72		55.98	40.56	179.25		
Thr[47]	113.57	8.65	4.42	4.61	59.96	71.92	176.19	$H^{\gamma 2}$ 1.21	$C^{\gamma 2}$ 21.76
Ser[48]	110.13	8.82	4.62	4.00, 3.86	62.74	60.28	176.31		
Gln[49]	120.51	7.23	3.77	2.32, 2.02	59.37	28.82	178.24	$H^{\epsilon 2}$ 7.25, 6.69	$N^{\epsilon 2}$ 110.76
Gln[50]	120.45	7.67	4.01	2.41, 2.04	59.06	29.55	179.15	$H^{\epsilon 2}$ 7.56, 6.72	$N^{\epsilon 2}$ 111.38
Gly[51]	107.01	8.27	3.73, 3.44		47.24		174.68		
Leu[52]	120.45	8.11	3.92		57.79	40.65	178.74		
Ala[53]	119.20	7.74	4.03	1.50	54.67	18.22	179.56		
Arg[54]	112.95	7.20	4.34	1.93	56.71	31.25	178.64		
Leu[55]	118.88	7.52	4.18		59.19	39.82	173.83		
Pro[56]			3.55		67.06	30.83	179.66		
Ala[57]	117.90	8.01	4.10	1.42	55.10	18.53	180.57		
Leu[58]	117.63	7.49	4.14	2.22	52.68		180.27		
Leu[59]	117.95	8.44	3.79		57.68	40.60	179.86		
Lys[60]	118.88	7.60	4.01	1.83, 1.44	58.71	32.76	178.03		
Gln[61]	115.13	7.97	3.92	1.65, 1.50	58.02	29.37	177.63	H^{γ} 2.08, 1.90	
His[62]	110.13	7.92	4.70	3.02, 2.85	55.94	30.60	173.26	$H^{\delta 2}$ 6.85	
Gln[63]	113.88	7.89	4.13	2.12, 1.97	55.90	28.20	172.71	H^{γ} 2.27, 2.27; $H^{\epsilon 2}$ 7.49, 6.79	$N^{\epsilon 2}$ 111.38
Pro[64]			4.32		62.54		175.09		
Arg[65]	118.57	8.34	4.16	0.89	56.49	29.83	176.31	H^{γ} 0.48; H^{δ} 2.78, 2.51; H^{ϵ} 6.87	N^{ϵ} 84.33
Trp[66]	115.13	7.59	5.60	3.15, 3.02	56.50	32.59	175.39	$H^{\epsilon 1}$ 10.17; $H^{\delta 1}$ 7.13; $H^{\eta 2}$ 7.14; $H^{\zeta 3}$ 6.95; $H^{\zeta 2}$ 7.28	$N^{\epsilon 1}$ 129.20
Val[67]	120.76	8.47	5.17	1.81	60.08	34.98	173.97	H^{γ} 0.70, 0.66	C^{γ} 21.44, 19.63
Leu[68]	128.89	9.30	5.09	2.03, 1.27	53.99	43.61	173.77		
Val[69]	127.95	9.16	4.29	2.23	62.49	32.96	174.07	H^{γ} 0.93, 0.82	C^{γ} 20.93, 20.31
Glu[70]	134.20	8.83	5.01	2.30	54.80		171.74		
Leu[71]	117.01	8.66	4.62		53.79	50.41	177.35		
Gly[72]	106.07	8.74	4.80, 3.86		44.07		176.31		
Gly[73]	111.07	9.43	4.14, 3.69		47.14		174.27		
Asn[74]	113.26	8.39	4.95	2.83, 2.50	53.53	38.90	175.49	$H^{\delta 2}$ 7.39, 6.86	$N^{\delta 2}$ 110.76
Asp[75]	122.32	7.89	3.82	3.09	58.09	38.67	178.54		
Gly[76]	108.57	9.02	3.64, 3.05		46.96		177.22		
Leu[77]	121.70	7.47	4.06	1.80, 1.63	56.80	41.81	178.34		
Arg[78]	114.51	7.09	4.36	2.07, 1.58	53.70	29.46	176.10	H^{δ} 3.22; H^{ϵ} 7.88	N^{ϵ} 84.06
Gly[79]	105.76	7.63	4.04, 3.67		45.56	177.93			
Phe[80]	118.57	7.79	4.73	2.94, 2.89	56.61	38.38	178.03	H^{δ} 7.16; H^{ϵ} 7.01; H^{ζ} 6.50	
Gln[81]	124.51	9.39	4.78	1.92, 1.92	54.37	27.48	176.25	H^{γ} 2.59, 2.59; $H^{\epsilon 2}$ 7.50, 6.91	$N^{\epsilon 2}$ 112.32
Pro[82]			4.00		66.80	31.43	177.02		

Table 2. Continued

Residue	^{15}N	NH	H$^{\alpha}$	H$^{\beta}$	^{13}C$^{\alpha}$	^{13}C$^{\beta}$	^{13}CO	Other ^{1}H	Other ^{15}N, ^{13}C
Gln^{83}	113.57	8.99	4.07	2.03, 2.03	59.10	27.55	176.92	H$^{\gamma}$ 2.40, 2.40; H$^{\epsilon 2}$ 7.55, 6.72	N$^{\epsilon 2}$ 111.07
Gln^{84}	118.57	7.61	4.18	2.27, 2.05	58.75	28.44	179.05	H$^{\gamma}$ 2.41, 2.41; H$^{\epsilon 2}$ 7.68, 7.32	N$^{\epsilon 2}$ 113.26
Thr^{85}	120.45	7.85	3.82	4.16	68.32		175.49	H$^{\gamma 2}$ 1.23	C$^{\gamma 2}$ 21.21
Glu^{86}	119.88	8.71	3.50		60.48	28.97	177.04		
Gln^{87}	116.13	7.84	4.01	2.21, 2.08	58.44	27.98	178.45	H$^{\gamma}$ 2.49, 2.49; H$^{\epsilon 2}$ 7.40, 6.87	N$^{\epsilon 2}$ 111.07
Thr^{88}	117.32	8.11	3.89	4.23	66.68	67.91		H$^{\gamma 2}$ 1.08	C$^{\gamma 2}$ 21.76
Leu^{89}	120.13	8.65	3.82		57.89	42.46	178.54		
Arg^{90}	118.88	8.50	3.58	2.02	60.50	29.70	177.22		
Gln^{91}	117.01	7.46	3.97	2.32, 2.13	58.75	29.00	178.44	H$^{\gamma}$ 2.44; H$^{\epsilon 2}$ 7.14, 6.89	N$^{\epsilon 2}$ 111.38
Ile^{92}	119.51	8.04	3.40	2.14	65.97	37.61	177.12	H$^{\gamma 2}$ 0.72	C$^{\gamma 2}$ 16.41
Leu^{93}	118.57	8.19	3.42	2.24, 1.95	58.88	39.58	178.64	H$^{\gamma}$ 1.05; H$^{\delta}$ 0.26	
Gln^{94}	117.01	8.21	3.97	2.30, 2.07	59.02		178.95	H$^{\gamma}$ 2.62, 2.51; H$^{\epsilon 2}$ 7.39, 6.82	N$^{\epsilon 2}$ 109.82
Asp^{95}	118.88	8.45	4.26	3.12	55.40	36.69	178.13		
Val^{96}	119.20	8.27	3.36	2.04	67.53	31.69	177.42	H$^{\gamma}$ 0.97, 0.61	C$^{\gamma}$ 22.31
Lys^{97}	117.01	8.04	4.45	1.95, 1.53	59.50	31.59	180.98		
Ala^{98}	121.38	8.39	4.22	1.50	54.14	18.32	177.83		
Ala^{99}	119.51	7.30	4.45	1.44	51.10	19.20	176.10		
Asn^{100}	112.95	8.14	4.29	3.06, 2.81	54.24	36.55	173.06	H$^{\delta 2}$ 7.48, 6.77	N$^{\delta 2}$ 112.01
Ala^{101}	118.26	7.42	4.58	0.90	49.22	21.65	175.49		
Glu^{102}	120.45	8.18	5.04	2.10, 2.10	51.90	30.87	174.49	H$^{\gamma}$ 2.51, 2.34	
Pro^{103}			5.12		61.08	31.83	174.27		
Leu^{104}	119.20	8.91	4.60	1.82, 1.12	53.37	43.90	175.09		
Leu^{105}	124.51	8.45	4.82	2.08, 1.12	53.90	45.49	173.87		
Met^{106}	124.51	8.04	4.80	2.06	54.71	35.14	174.03		
Gln^{107}	125.82	8.59	3.06		56.54		172.26		
Ile^{108}			3.01	1.53	56.53	36.64		H$^{\gamma 2}$ 0.54	C$^{\gamma 2}$ 16.81
Arg^{109}			3.50		65.13	31.03	177.12		
Leu^{110}	114.51	7.21	4.25	1.67, 1.27	55.74	42.43	178.03		
Pro^{111}			4.26		62.21	31.56	176.61		
Ala^{112}	122.01	8.10	3.94	1.19	53.21	18.59	178.13		
Asn^{113}	112.32	8.18	4.41	2.64, 2.62	54.00	37.01	175.60	H$^{\delta 2}$ 7.43, 6.57	N$^{\delta 2}$ 111.38
Tyr^{114}	119.82	7.68	4.46	3.05, 2.85	57.47	37.74	176.81	H$^{\delta}$ 6.80; H$^{\epsilon}$ 6.58	
Gly^{115}	106.70	7.92	4.37, 3.86		44.43		173.66		
Arg^{116}	120.45	8.43	4.06	1.85, 1.63	59.57	30.51	178.03	H$^{\delta}$ 3.17; H$^{\epsilon}$ 7.32	N$^{\epsilon}$ 83.91
Arg^{117}	117.95	8.48	4.61	1.73	59.41	29.43	179.35	H$^{\delta}$ 3.22; H$^{\epsilon}$ 7.30	N$^{\epsilon}$ 84.10
Tyr^{118}	120.13	8.45	4.09	3.07, 2.86	61.68	38.22	176.81	H$^{\delta}$ 7.25; H$^{\epsilon}$ 6.46	
Asn^{119}	117.32	8.52	4.29	2.97, 2.83	56.95	38.05	178.74		
Glu^{120}	119.20	8.65	4.08	2.12, 2.12	55.16	32.06	178.64	H$^{\gamma}$ 2.55, 2.38	
Ala^{121}	121.01	7.54	4.13	1.44	54.51	18.25	179.66		
Phe^{122}	117.63	9.09	3.92	2.86, 2.36	61.12	40.37	177.02	H$^{\delta}$ 6.83; H$^{\epsilon}$ 7.12	
Ser^{123}	111.70	8.41	4.11	1.99	62.52	57.22	176.51		
Ala^{124}	118.88	7.20	4.16	1.49	53.38	18.36	178.64		
Ile^{125}	116.38	7.46	3.32	1.35	65.16	38.12	176.51	H$^{\gamma 2}$ 0.54; H$^{\delta 1}$ 0.30	C$^{\gamma 2}$ 16.75; C$^{\delta 1}$ 14.43
Tyr^{126}	113.26	6.70	3.84	3.13, 2.69	64.21	35.78	174.29	H$^{\delta}$ 6.92; H$^{\epsilon}$ 6.50	

Table 2. Continued

Residue	^{15}N	NH	H^{α}	H^{β}	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	^{13}CO	Other ^{1}H	Other ^{15}N, ^{13}C
Pro[127]			4.15		65.33	30.30	179.15		
Lys[128]	117.01	6.84	3.99	1.85	59.76	32.71	179.35		
Leu[129]	118.88	8.30	4.18	1.74, 1.14	57.38	42.18	176.71		
Ala[130]	120.76	8.33	4.04	1.61	55.49	18.25	179.86		
Lys[131]	115.45	7.65	4.23	1.94, 1.61	58.48	32.51	179.35	H^{γ} 1.51	
Glu[132]	119.82	8.27	3.83	1.61, 1.61	59.07	29.73	178.03	H^{γ} 2.20, 2.20	
Phe[133]	112.01	8.07	4.30	3.15, 2.77	59.00	40.37	174.27	H^{δ} 7.43; H^{ϵ} 7.26; H^{ζ} 7.32	
Asp[134]	121.70	7.96	4.46	3.29, 2.55	54.35	37.31	174.68		
Val[135]	112.01	8.39	4.99	2.56	57.21	31.84	173.03	H^{γ} 0.99, 0.80	C^{γ} 21.72, 18.89
Pro[136]			4.23		63.54	32.79	174.07		
Leu[137]	122.70	8.01	5.29	1.89	52.17	43.61	176.88	1.07	
Leu[138]	129.57	9.23	4.82	1.74	51.94	42.58	173.58	1.37	
Pro[139]			4.55		61.81	31.89	174.78		
Phe[140]	117.70	8.65	4.83	3.04, 3.04	54.52		179.21	H^{δ} 7.29; H^{ϵ} 6.82; H^{ζ} 6.54	
Phe[141]	125.82	6.86	4.35	3.18, 3.00			175.27	H^{δ} 6.99; H^{ϵ} 7.36; H^{ζ} 6.57	
Met[142]			4.34		53.17	28.50	176.61		
Glu[143]	115.76	7.02	4.00	2.14, 1.98	58.84	28.17	178.13		
Glu[144]	112.63	7.16	4.16	2.10, 1.70	56.87	29.24	176.61		
Val[145]	118.57	6.90	3.79	1.80	64.65	32.03	179.05	H^{γ} 0.93, 0.84	C^{γ} 22.85
Tyr[146]	113.95	7.83	4.49	3.29	58.54	38.45	177.01	H^{δ} 7.21; H^{ϵ} 6.55	
Leu[147]	122.38	8.10	4.18	1.57, 1.50	55.44	41.98	177.37	0.79	
Lys[148]	118.95	7.99	4.11		56.16	29.41	176.08		
Pro[149]					65.47	31.63	178.95		
Gln[150]	114.82	9.09	4.26	1.96	57.37	26.64	176.51	$H^{\epsilon 2}$ 7.21, 6.58	$N^{\epsilon 2}$ 110.76
Trp[151]	122.63	8.52	5.31	3.42, 3.41	54.34	31.36	175.29	$H^{\epsilon 1}$ 9.90; $H^{\delta 1}$ 6.82; $H^{\epsilon 3}$ 7.22; $H^{\zeta 3}$ 5.96; $H^{\eta 2}$ 6.66; $H^{\zeta 2}$ 7.03	$N^{\epsilon 1}$ 125.76
Met[152]	117.63	7.36	5.23	2.03	51.90	32.03	177.83		
Gln[153]	117.32	8.66	4.59	2.20, 1.35	56.62	28.30	177.32	$H^{\epsilon 2}$ 7.21, 6.58	$N^{\epsilon 2}$ 111.07
Asp[154]	120.13	8.98	4.31	2.84	55.75	39.17	175.80		
Asp[155]	114.82	7.27	4.41	3.04, 2.88	53.49	39.97	177.12		
Gly[156]	107.63	8.16	4.02, 3.66		46.09		172.95		
Ile[157]	117.63	8.47	4.03	0.93	61.36	41.88	175.09	$H^{\gamma 1}$ 1.20; $H^{\gamma 2}$ 0.65; $H^{\delta 1}$ 0.15	$C^{\gamma 1}$ 26.19; $C^{\gamma 2}$ 15.45; $C^{\delta 1}$ 12.68
His[158]	115.76	8.69	4.76	3.26, 2.70	55.47	29.30	171.88	$H^{\delta 2}$ 6.61; $H^{\epsilon 1}$ 7.50	
Pro[159]			3.50		64.74	32.03	177.12		
Asn[160]	113.95	7.82	4.25		58.54	38.45	178.03		
Arg[161]	112.01	7.38	3.55	1.12	61.12	31.25	175.49	H^{γ} 0.06, −0.34; H^{ϵ} 6.60	N^{ϵ} 83.98
Asp[162]	118.26	7.46	4.56	2.67, 2.55	53.03	38.63	174.99		
Ala[163]									
Gln[164]	120.19	8.41	4.22		54.45	40.39	176.20		
Pro[165]			3.62		66.40	30.50	178.54		
Phe[166]	117.32	7.17	4.05	2.67, 2.34	60.89	38.33	177.53	H^{δ} 7.00; H^{ϵ} 6.58; H^{ζ} 6.72	
Ile[167]	119.20	7.79	2.78	1.14	65.26	38.18	177.42	$H^{\gamma 2}$ 0.58; $H^{\delta 1}$ – 1.17	$C^{\gamma 2}$ 16.30; $C^{\delta 1}$ 14.09

Table 2. Continued

Residue	^{15}N	NH	H^{α}	H^{β}	$^{13}C^{\alpha}$	$^{13}C^{\beta}$	^{13}CO	Other ^{1}H	Other ^{15}N, ^{13}C
Ala[168]	119.82	8.44	4.37	1.67	55.90	18.27	178.44		
Asp[169]	117.32	7.70	4.14	2.63, 2.56	57.02	40.24	177.63		
Trp[170]	122.63	8.43	3.73	2.89	61.76	28.97	179.46	$H^{\epsilon 1}$ 10.13; $H^{\delta 1}$ 7.14; $H^{\eta 2}$ 7.11; $H^{\zeta 3}$ 6.83; $H^{\zeta 2}$ 7.40	$N^{\epsilon 1}$ 129.51
Met[171]	117.01	8.95	4.07		55.54	29.41	178.44		
Ala[172]	122.01	8.34	3.76	1.29	55.49	16.79	179.66		
Lys[173]	115.76	7.35	3.90	1.65	58.93	32.00	179.35		
Gln[174]	115.45	7.53	3.67	1.80, 1.46	56.92	27.76	177.93	H^{γ} 2.13, 1.69; $H^{\epsilon 2}$ 7.20, 6.63	$N^{\epsilon 2}$ 109.20
Leu[175]	112.63	8.34	4.11	1.99, 1.31	55.84	43.25	178.24		
Gln[176]	120.76	8.05	3.91		61.19		173.94		
Pro[177]			4.50		61.68	32.22	177.12		
Leu[178]	112.95	6.86	4.09	1.56	51.09	39.63	176.92		
Val[179]	119.82	8.79	3.58	1.18	59.87	30.44	178.74	H^{γ} 0.08, −0.31	C^{γ} 19.06, 18.56
Asn[180]	119.82	7.96	4.35	2.90, 2.80	56.14	39.48	176.31		
His[181]	117.63	7.45	2.81	1.38	55.93	28.86	178.27	$H^{\delta 2}$ 6.87; $H^{\epsilon 1}$ 7.73	
Asp[182]									
Ser[183]									
His[184]									
His[185]									
His[186]									
His[187]									
His[188]									
His[189]									

constant time mode. This experiment also had the advantage that the cross peaks could be edited according to the number of coupled partners as either positive peaks for those ^{13}C nuclei coupled to an odd number of spins or negative peaks for those coupled to an even number of spins. Thus, the threonine C^{β}-$C^{\beta}H$ and glycine C^{α}-$C^{\alpha}H$ cross peaks had intensities opposite in sign to those of the C^{α}-$C^{\alpha}H$ cross peaks of all the other amino acids, making them readily distinguishable. The HNCACB and CBCA(CO)NH experiments provided both sequential connectivity and information on the residue type from the unique $^{13}C^{\alpha}/^{13}C^{\beta}$ chemical shift patterns of some residues (Grzesiek and Bax, 1993), especially for alanine, glycine, serine, and threonine residues. We identified 18 out of 19 alanine, 10 out of 10 glycine, 8 out of 11 serine, and 5 out of 5 threonine spin systems. In the later stage of assignment we also used $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift data to identify valine, isoleucine, and proline residues whose $^{13}C^{\alpha}$ nuclei were expected to resonate at lower fields (usually greater than 60 ppm).

Backbone assignment

Backbone assignments were accomplished based primarily on the following six 3D triple-resonance experiments: HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HN(CA)CO, and HNCO. Figure 2 shows the assignment of backbone $^{13}C^{\alpha}$ resonances for residues Asp[3] to Asp[10] based on the HNCA and HN(CO)CA experiments. The HN(CO)CA spectrum identified the interresidue connectivity, while the HNCA spectrum identified both the intra- and interresidue connectivities and hence could detect degenerate resonances from adjacent residues. Similarly, the sequence connectivities based on C^{α} and C^{β} chemical shifts were obtained based on the CBCA(CO)NH and HNCACB spectra. The side-chain amide ^{15}N resonances of asparagine and glutamine, which are expected to resonate in the range of 107–115 ppm, were also identified on a 3D CBCA(CO)NH spectrum. The $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts of all proline residues, except the C^{β} of Pro[64], were assigned based on the CBCA(CO)NH experiment. The HNCO and

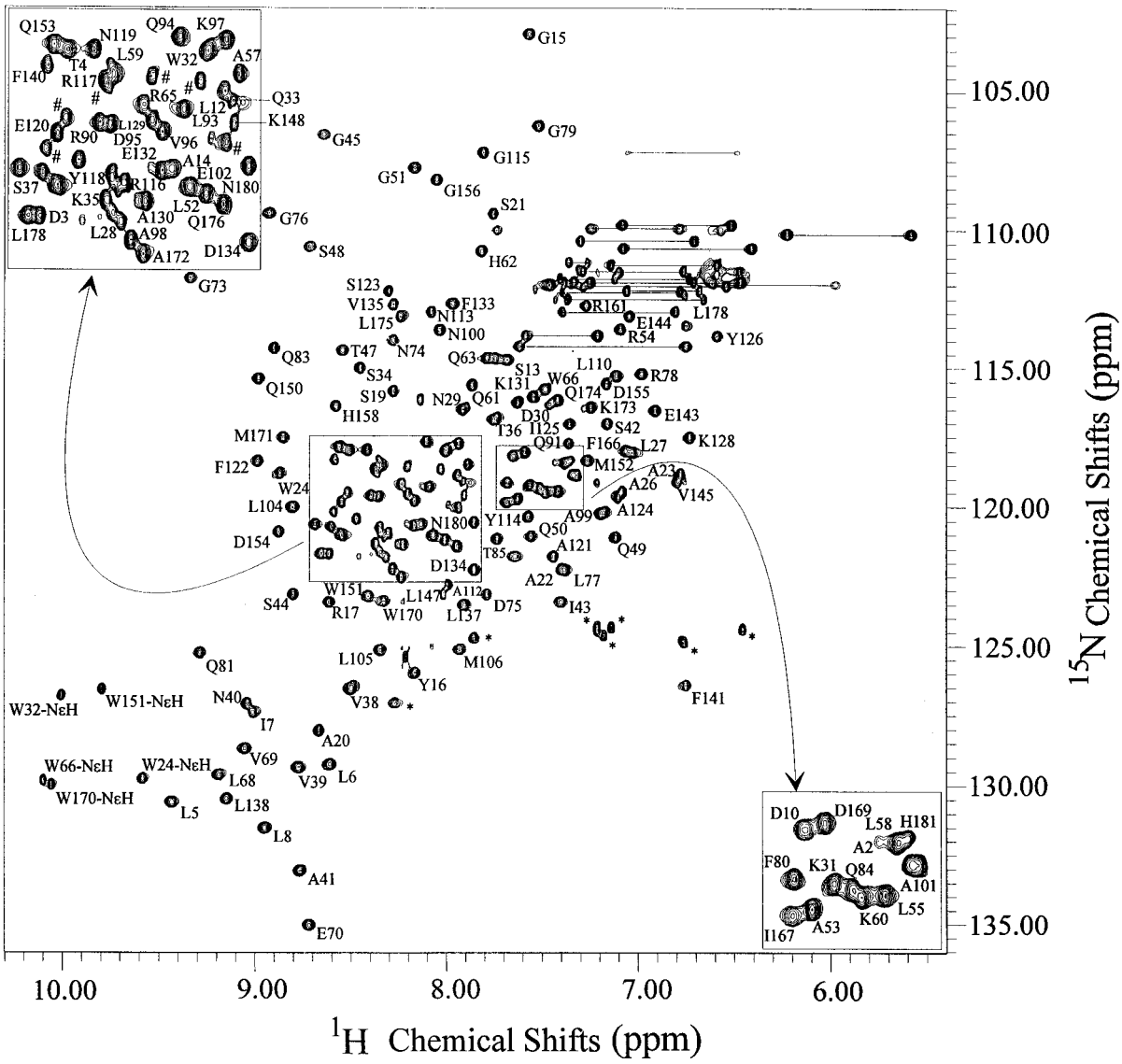

Figure 1. 2D ^{1}H-^{15}N-HSQC spectrum of uniformly ^{15}N-enriched *E. coli* thioesterase/protease I. The sample concentration was 3 mM, pH 3.5, 310 K. Assignments of backbone amide protons and ^{15}N cross peaks are also shown. Two regions are expanded for clarity as indicated by arrows. The tryptophan ring N$^{\epsilon}$H's and ^{15}N's are also shown. The horizontal lines connect the pairs of the side-chain NH_2 protons in asparagine and glutamine residues. The cross peaks marked by asterisks are the folded arginine side-chain N$^{\epsilon}$H protons. The cross peaks labeled with '#' are from the C-terminal histidine tags.

HN(CA)CO experiments were used to confirm the sequential connectivity and to obtain ^{13}CO chemical shifts. After careful analysis of the six triple-resonance experiments, we assigned the backbone resonances of all but 16 residues. To confirm the assignments and to further assign the unassigned resonances, we prepared three protein samples, each labeled with ^{15}N at one of the specific amino acids: alanine, leucine, and phenylalanine. We observed cross peaks in the respective 2D ^{15}N-HSQC spectra for 6 out of 6 phenylalanines, 18 out of 19 alanines, and 24 out of 24 leucines. These confirmed all of the previous assignments and provided eight additional assignments (Ala^{2}, Ala^{57}, Phe^{140}, Phe^{141}, Leu^{138}, Leu^{147}, Ala^{168}, and Leu^{178}). The assignments of Phe^{140} and Phe^{141} in the sequence Pro^{139}-Phe^{140}-Phe^{141}-Met^{142} were difficult prior to obtaining the [^{15}N-Phe] HSQC spectrum since Met^{142} could not be identified and position

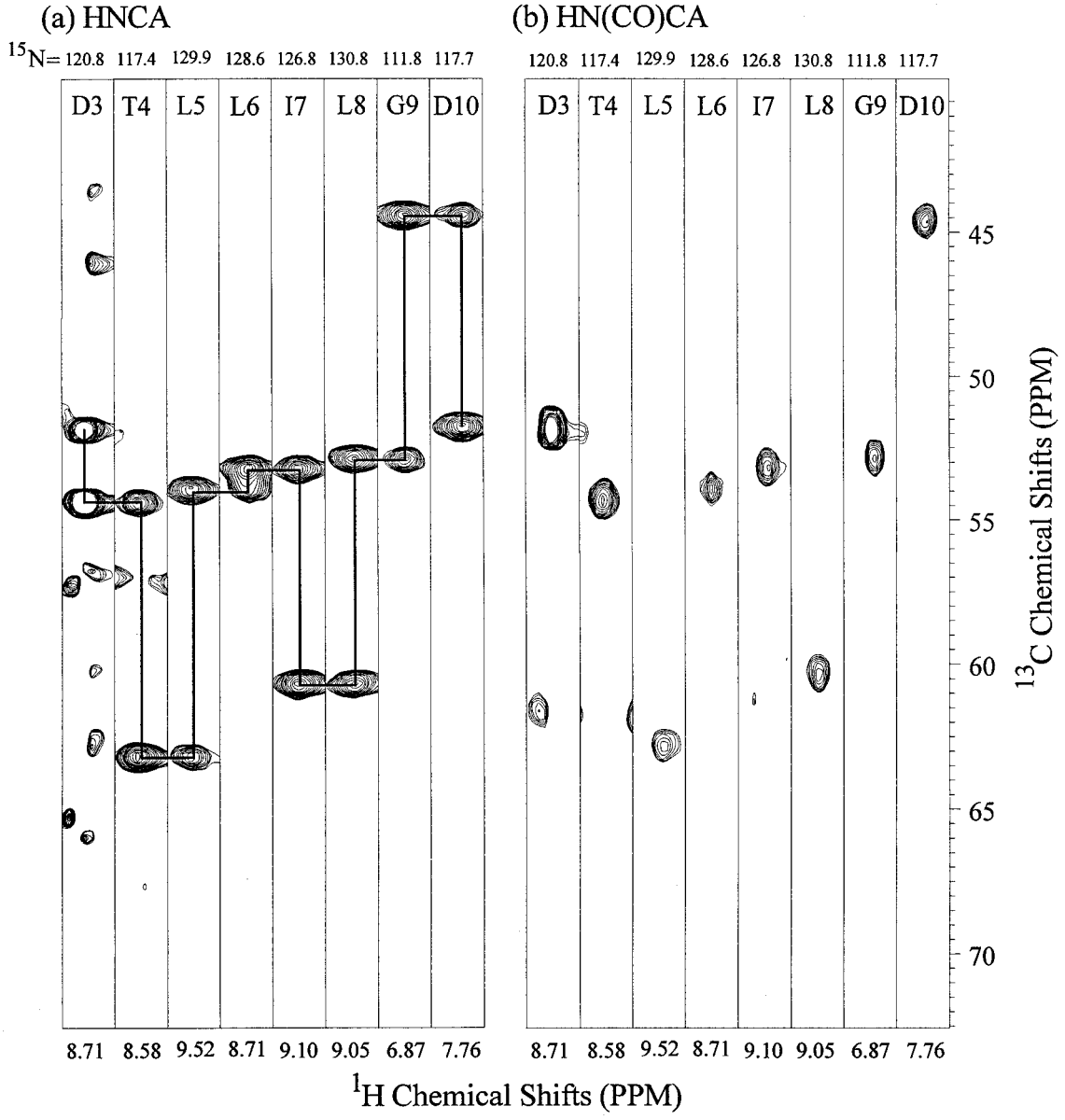

Figure 2. Strip plot of data extracted from (a) 3D HNCA and (b) HN(CO)CA spectra of thioesterase/protease I to show the sequential backbone resonance assignment for residues Asp^{3} to Asp^{10}.

139 is a proline residue. However, in the [^{15}N-Phe] 2D HSQC spectrum, we observed six cross peaks. Four of them had already been previously assigned to phenylalanines. The two unassigned cross peaks were then readily assigned to Phe^{140} and Phe^{141} based on HNCACB and 3D NOESY-HSQC experiments. The overlay of four 2D ^{15}N-HSQC experiments for three specifically ^{15}N-labeled proteins and one uniformly ^{15}N-labeled protein is shown in Figure 3. The sequential assignments were further verified with intra- and interresidue NOEs obtained from analyzing the 3D ^{15}N-NOESY-HSQC spectrum from which we were able to assign additional H^{α} and H^{β} resonances. The 3D ^{15}N-NOESY-HSQC experiment was also very useful in reconfirming the resonance assignments of asparagine and glutamine residues since all the assigned NH_2 groups were observable in the 3D ^{15}N-NOESY-HSQC spectrum. Table 2 summarizes the chemical shift assignments of backbone nuclei for *E. coli* thioesterase/protease I. In the present study we

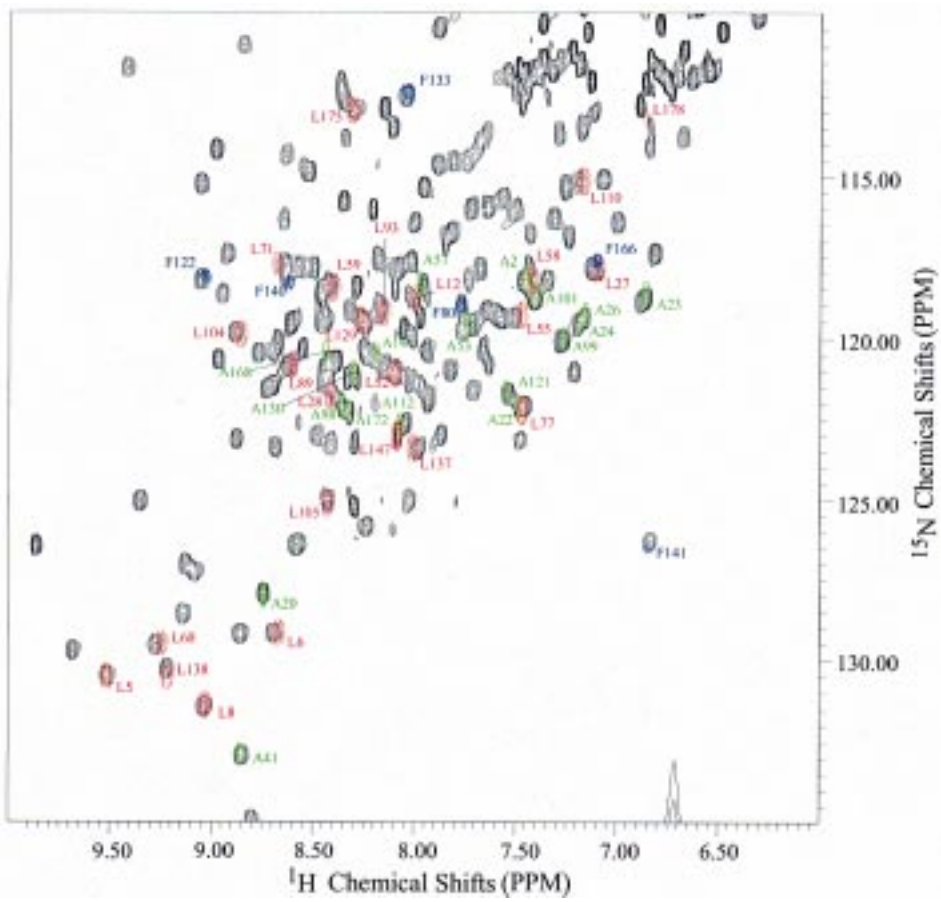

Figure 3. Overlay of four 2D ^{15}N-HSQC spectra acquired with four different *E. coli* thioesterase/protease I samples differing in ^{15}N-labeling positions, but in the same solvent (pH 3.5 in 90% H_2O) and under similar NMR conditions (310 K). Samples employed were: black – uniformly ^{15}N-labeled; blue – [^{15}N]-Phe-labeled; red – [^{15}N]-Leu-labeled; green – [^{15}N]-Ala-labeled.

were still unable to obtain complete backbone resonance assignments of six residues (Met^{1}, Ser^{11}, Ile^{108}, Arg^{109}, Met^{142}, and Ala^{163}). All of the incompletely assigned residues are in the loop region, as described below. Some very strong cross peaks were observed in the ^{15}N-HSQC spectrum which were assigned to the six histidine tags in the C-terminus. No attempt was made to assign these resonances.

Side-chain assignment

Side-chain ^{1}H and ^{13}C resonances were assigned using 3D HCCH-TOCSY (Bax et al., 1990) and 3D C(CO)NH (Grzesiek et al., 1993) experiments, respectively. Since the ^{13}C chemical shifts of the methyl groups of alanine, valine, isoleucine, and threonine have most upfield chemical shifts, we also used this information to reconfirm our assignment for these residues. Figure 4 shows the side-chain assignment of seven isoleucines, extracted from the HCCH-TOCSY spectrum. In addition, Ile^{108}, which was previously unassigned, was identified by exclusion. The C(CO)NH experiment was used to confirm the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shift assignments and to provide assignments for the remaining side-chain carbons. The 2D TOCSY and NOESY spectra acquired in D_2O were used for assigning aromatic side chains. The connectivities between the aromatic rings and the aliphatic protons of all tyrosines and phenylalanines were established from their respective $H^{\beta}H^{\delta}$ NOE cross peaks. Resonances for all the aromatic protons of the five tyrosines and the six phenylalanines were subsequently fully assigned based on the 2D TOCSY spectrum. The side-chain $N^{\epsilon}H$ resonances of the five tryptophans were identified in the 2D ^{15}N-HSQC spectrum and their sequence positions were assigned based on the NOEs of $N^{\epsilon}H/C^{\beta}H$ or $N^{\epsilon}H/C^{\alpha}H$. The assignments of

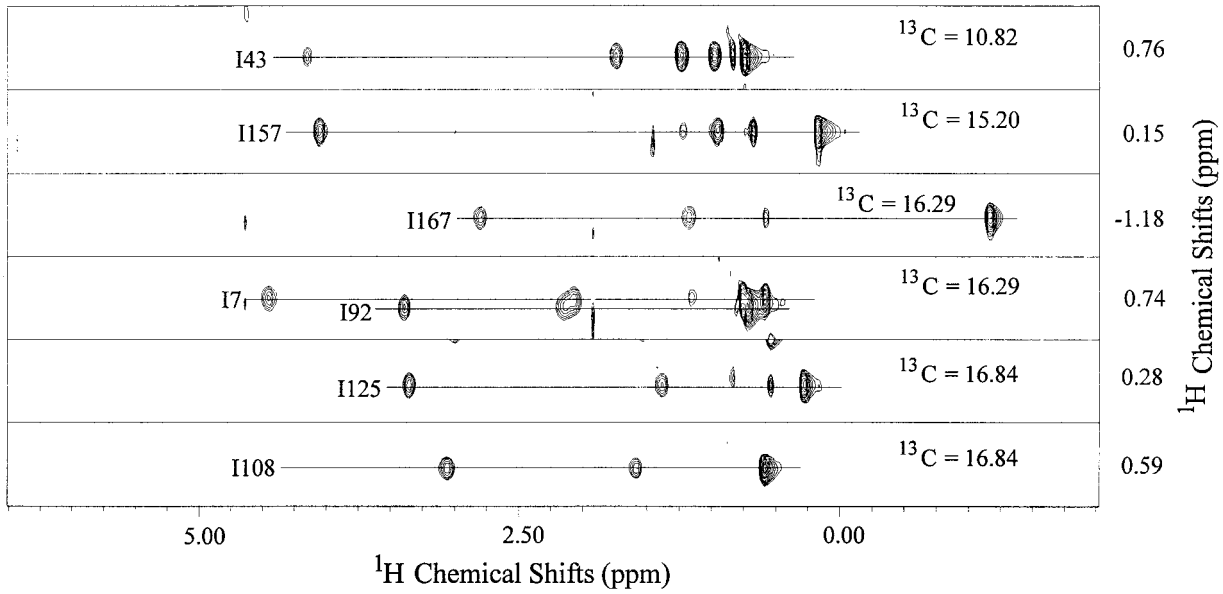

Figure 4. Strip plot extracted from a 3D HCCH-TOCSY spectrum obtained with uniformly ^{13}C-labeled thioesterase/protease I in D_2O, showing the side-chain proton assignments of all seven isoleucine residues.

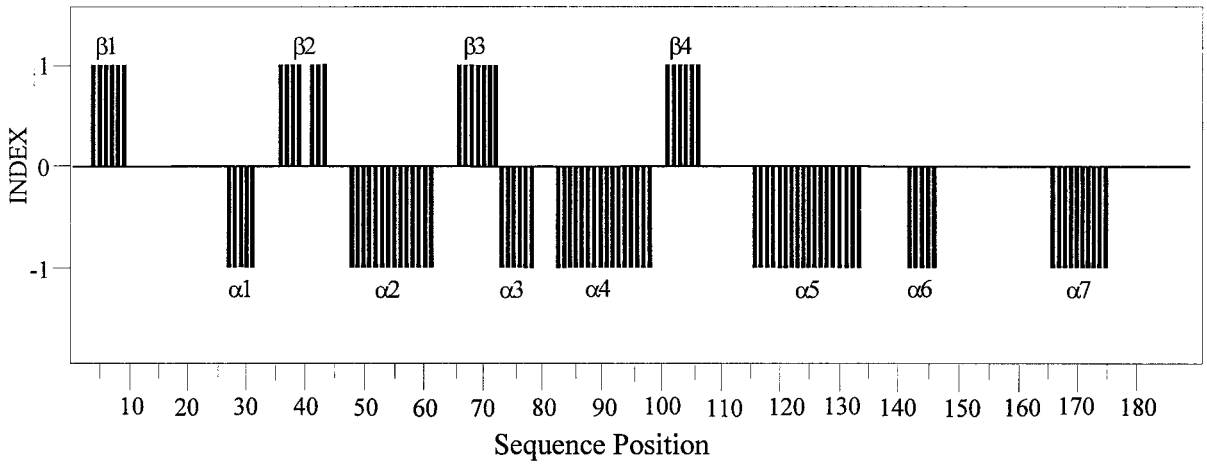

Figure 5. CSI consensus plot for *E. coli* thioesterase/protease I, determined using four nuclei ($^{1}H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO). The secondary structural motifs obtained from this program are summarized in the figure.

C^2H and C^7H for tryptophan residues were based on the relative intensities of the NOE cross peaks between these protons and the $N^{\epsilon}H$, where the $N^{\epsilon}H/C^2H$ NOE cross peaks were more intense than the $N^{\epsilon}H/C^7H$ cross peaks. The other tryptophan aromatic protons were readily assigned from their coupling to the C^7H protons which were identified from the TOCSY spectrum.

Secondary structure determination

The secondary structure of a protein can be established from a number of NMR-determined parameters, including chemical shift differences from random coil values (Dalgarno et al., 1983; Spera and Bax, 1991; Wishart et al., 1991), the magnitude of NH-$C^{\alpha}H$ coupling constants, and the patterns of short- and medium-range NOEs (Wüthrich, 1986). Using the nearly completed H^{α}, C^{α}, C^{β}, and CO resonance assignments, we first employed the consensus chemical shift index (CSI) (Wishart and Sykes, 1994) to identify the secondary structure of *E. coli* thioesterase/protease I. A residue was assigned to a helix structure only if at

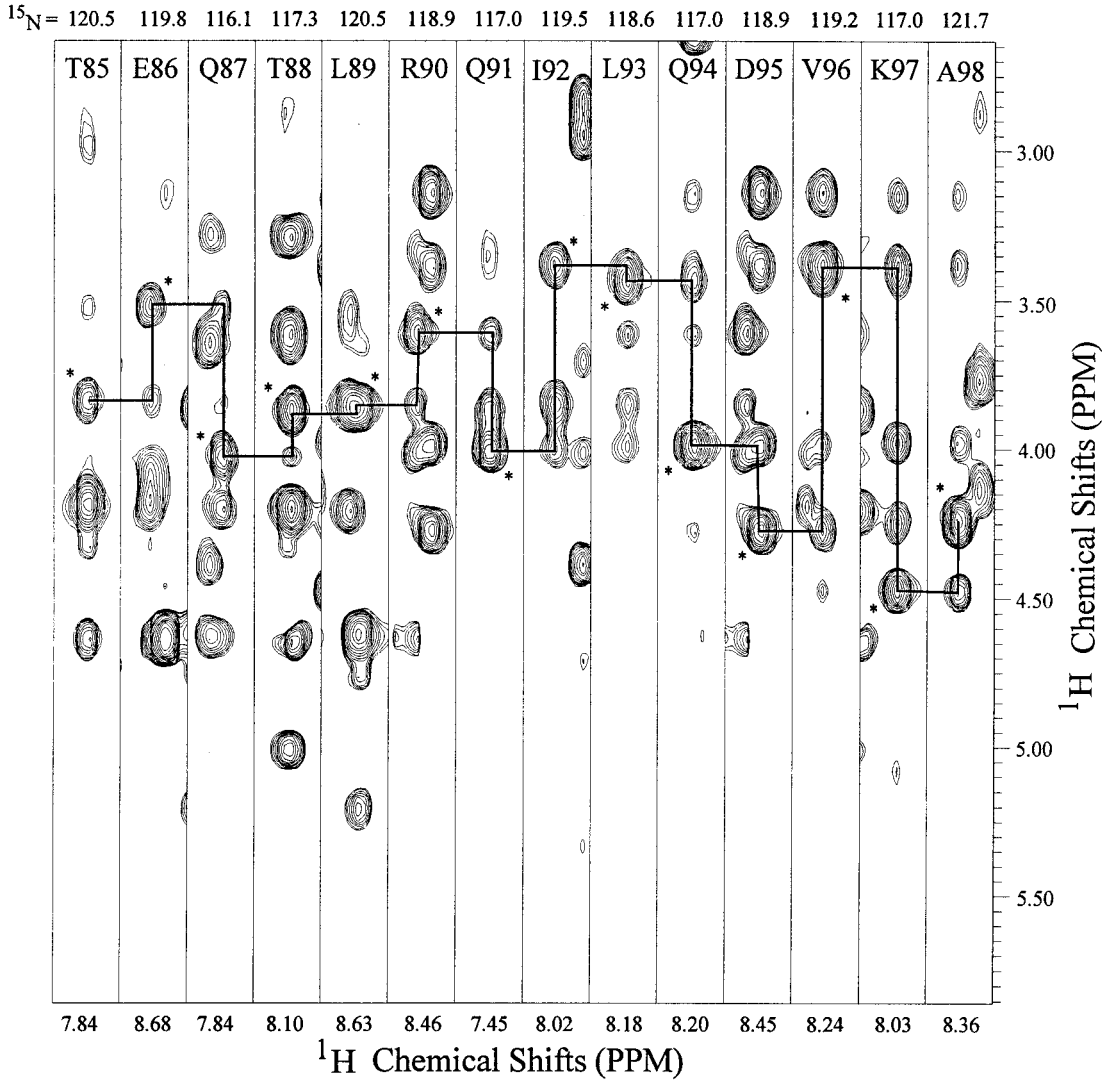

Figure 6. Slices of the fingerprint region of a 110 ms 3D ^{15}N-NOESY-HSQC spectrum illustrating the sequential connectivities from Thr^{85} to Ala^{98}. The intense intra-$d_{\alpha N}$(i, i) NOEs are marked.

least two out of the three H^{α}, C^{α}, and CO CSIs agreed with a helix structure. A residue was assigned as a β-strand structure only if at least three out of the four H^{α}, C^{α}, C^{β}, and CO CSIs indicated a β-strand. The consensus plot of CSI is shown in Figure 5, indicating that *E. coli* thioesterase I is composed of seven α-helices and four β-strands. The seven α-helices are located in the range of Leu^{27}–Lys^{31} (helix α1), Ser^{48}–Gln^{61} (helix α2), Gly^{73}–Arg^{78} (helix α3), Gln^{83}–Ala^{98} (helix α4), Arg^{116}–Phe^{133} (helix α5), Met^{142}–Tyr^{146} (helix α6), and Phe^{166}–Leu^{175} (helix α7). The four β-strands are located in the range of Thr^{4}–Gly^{9} (β1-strand), Thr^{36}–Ile^{43} (β2-strand), Trp^{66}–Gly^{72} (β3-strand), and Ala^{101}–Met^{106} (β4-strand). The secondary structures deduced from the CSI analysis were further verified with short- and medium-range NOEs, extracted from the 3D ^{15}N-NOESY-HSQC spectrum acquired with a mixing time of 110 ms (Figure 6). Through analysis of the NOE pattern in the fingerprint region of the 3D ^{15}N-NOESY-HSQC spectrum, we verified most of the secondary structure assignments based on the backbone side-chain sequential connectivities. Thus, the α-helix is characterized by a series of strong d_{NN} NOEs, relatively weak $d_{\alpha N}$, $d_{\alpha N}$(i, i+2), $d_{\alpha N}$(i, i+4) NOEs, and medium $d_{\alpha N}$(i, i+3) NOEs. Figure 6 shows a representative region of the sequential as-

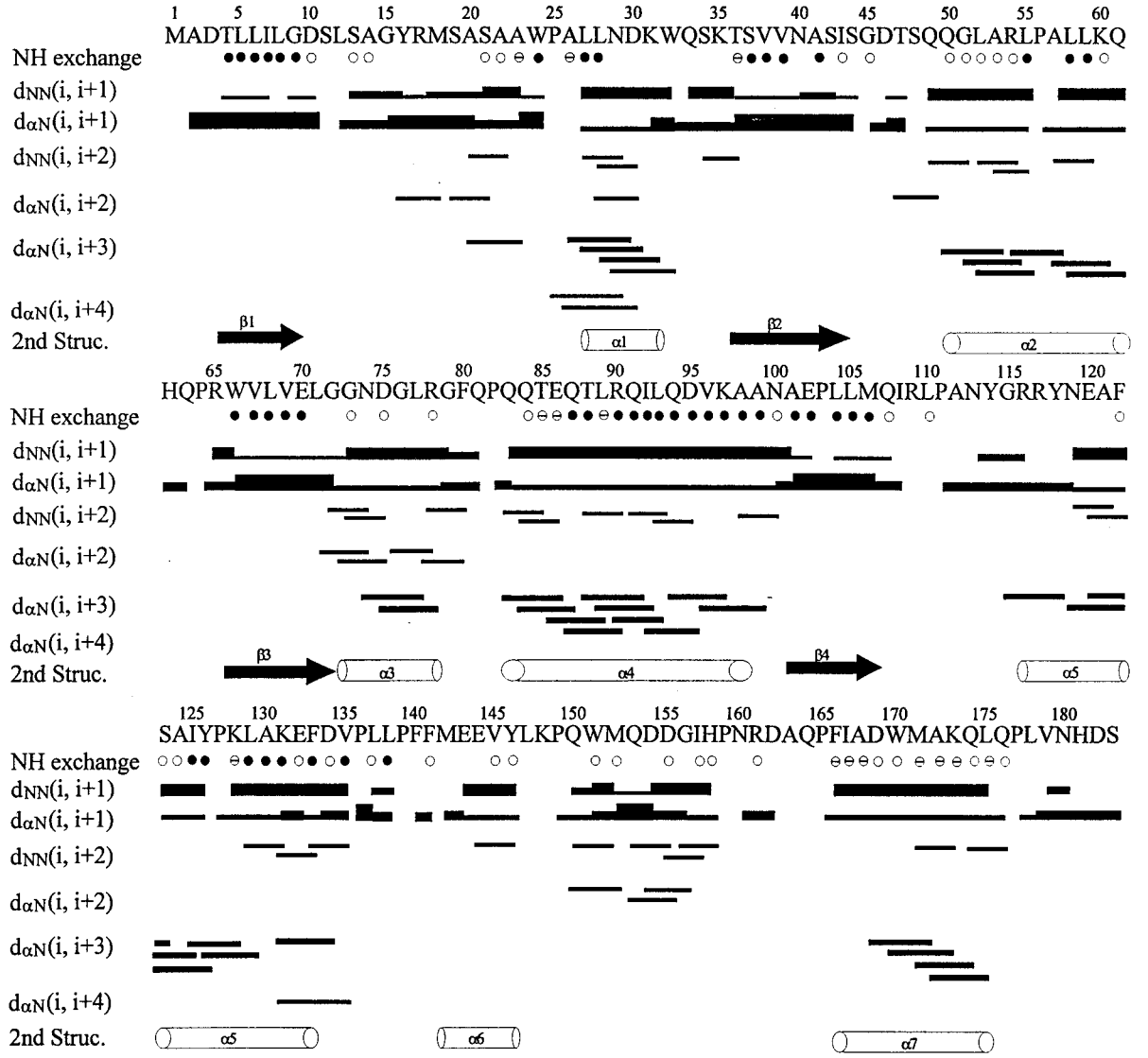

Figure 7. Sequence of *E. coli* thioesterase I, the amide proton exchange rates, and summary of the sequential and medium-range NOEs involving HN and H^{α} protons. Amide protons that remained observable in the 2D ^{15}N-HSQC spectrum after 6 days are labeled with filled circles (very slow-exchange protons), those that were observable after 2 days are marked with open circles with a central horizontal line (slow-exchange protons), those that were observable after 40 min are marked with open circles (medium-exchange protons), and those that were not observable upon the addition of D_2O (fast-exchange protons) are not marked. The NOE intensities are indicated by the thickness of the black bars, with thicker bars representing stronger NOEs.

signment from the 3D ^{15}N-NOESY-HSQC spectrum which exhibited the characteristics of α-helix NOE connectivities. We also checked the amide proton exchange rate data to confirm the secondary structures. Almost all the NH protons in the 98 β-strand regions exchanged slowly. Most of the NH protons in the long helices (α2, α4, α5, and α7) are exchanging slowly. NH protons in the short helices (α1, α3, and α6) exchange faster. This observation is consistent with the secondary structure assignments. A summary of short- and medium-range NOE connectivity patterns, backbone amide proton exchange rates, and the assigned secondary structures is given in Figure 7. From the cross-over NOE connectivity pattern and the slow amide proton exchange rates, as shown in Figure 8, the four β-strands were found to form a four-stranded parallel β-sheet structure.

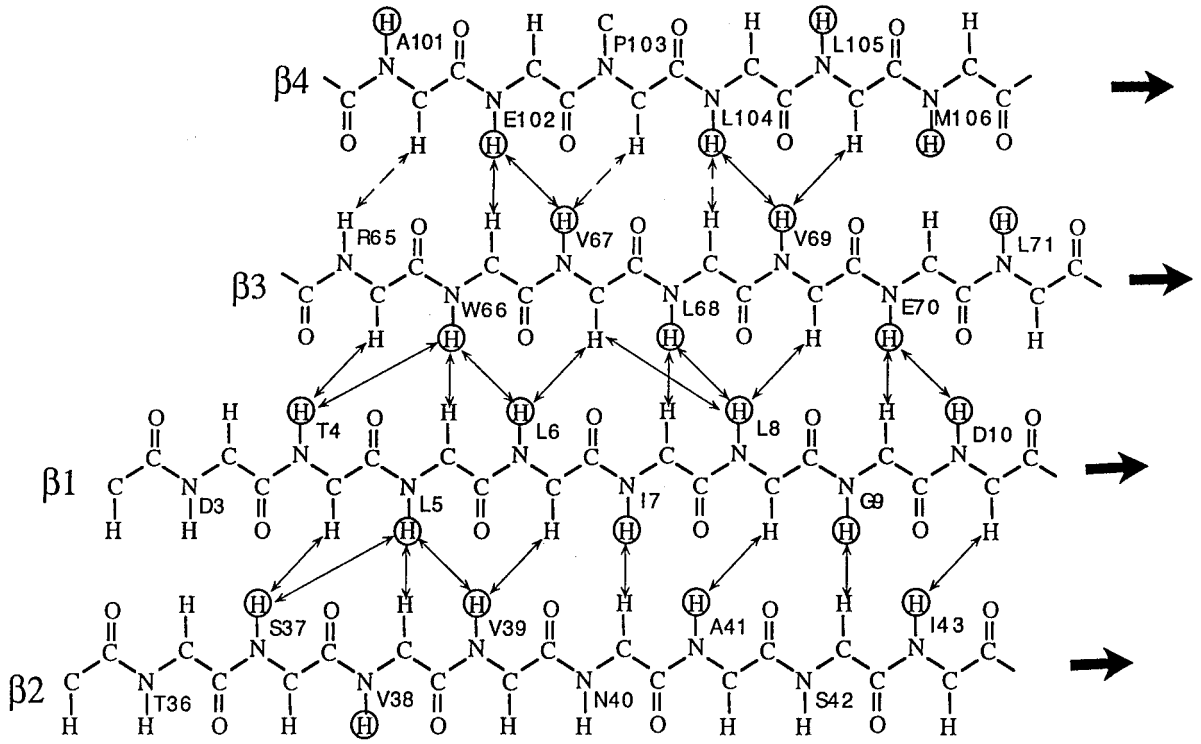

Figure 8. The proton NOE connectivity among the four β-strands identified in *E. coli* thioesterase/protease I. Arrows indicate the presence of NOE cross peaks between the two connected protons and the amide protons with medium and slow exchange rates are circled. Arrows with dashed lines indicate that NOEs could not be resolved due to resonance overlap. The connectivity pattern is consistent with the formation of a four-stranded parallel β-sheet secondary structure of *E. coli* thioesterase I.

Discussion and Conclusions

Although we have not determined the full three-dimensional structure of *E. coli* thioesterase/protease I, the secondary structure features identified in the present study revealed substantial clues concerning the structure–function relationship of this enzyme. The most striking feature of the secondary structure of *E. coli* thioesterase/protease I is its similarity to the core region of other lipases. Figure 9 shows the secondary structure topology of the *E. coli* thioesterase/protease I deduced from the present study. The presence of seven α-helices and four β-strands arranged in alternate order before α5 is typical of the α/β hydrolase fold. The four β-strands further form a parallel β-sheet. Thus, the topology of the four strands is $-1x$, $+2x$, $+1x$ (Richardson, 1981). When compared to the topology of the common α/β hydrolase proteins, such as carboxylpeptidase II from wheat (Breddam, 1986; Cooper and Bussey, 1989; Liao and Remington, 1990; Thomas et al., 1990), dienelactone hydrolase from *Pseudomonas* sp. B13 (Pathak et al., 1988; Pathak and Ollis, 1990), triacyl glycerol lipase from *R. miehei* (Brady et al., 1990), and human pancreas triacyl glycerol lipase (Winkler et al., 1990), the secondary folding of the *E. coli* thioesterase/protease I closely resembles the catalytic core of these lipolytic enzymes (strands 3–7 of the prototype α/β hydrolase fold) (Ollis et al., 1992). It is tempting to conclude that indeed the structure of *E. coli* thioesterase/protease I belongs to the α/β hydrolase superfamily, the same as the common lipases. However, there are substantial differences between the secondary structure of *E. coli* thioesterase/protease I and the common α/β hydrolase. One major difference is the number of β-strands present in the α/β hydrolase family, which is usually much higher, ranging from 8 in dienelactone hydrolase (MW 25 kDa) to 12 in acetylcholine esterase (MW 60 kDa). Another major difference is the location of the active site residues. In α/β hydrolase the nucleophile serine residue is located in a sharp turn in between strand five and helix six, corresponding to strand β3 and helix α3 in *E. coli* thioesterase/protease I. In comparison, the nucleophile serine is located at

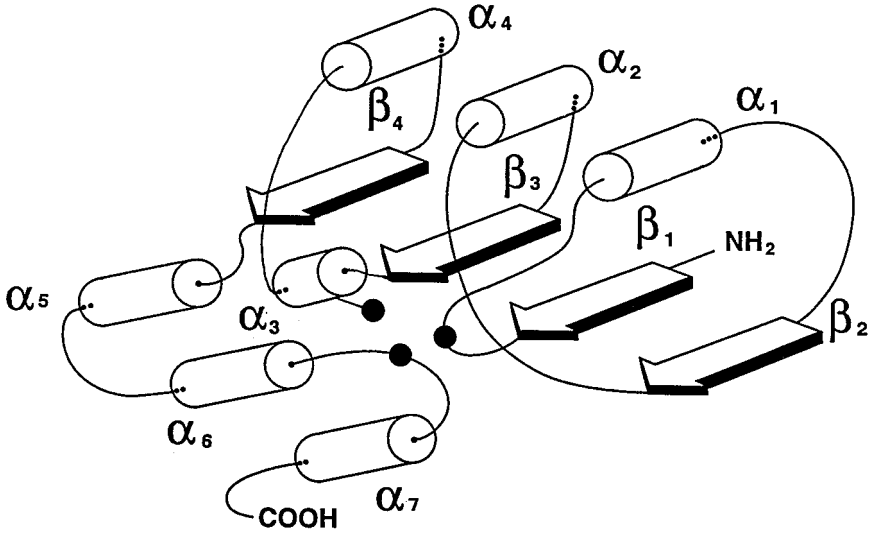

Figure 9. Topology of the four β-strands of *E. coli* thioesterase/protease I. In Richardson's notation (Richardson, 1981) (+/−)n<x>, n refers to the number of strands in the secondary structure between the current strand and the next strand in sequence, while + or − indicates whether the strand is to the left or to the right of the current strand. x indicates that there is a cross-over between the strands, so that they are parallel in direction. In this notation, the topology of *E. coli* thioesterase/protease I is −1x, +2x, +1x. Filled circles indicate the location of the putative catalytic triad residues.

the turn of the first β-strand in thioesterase/protease I. Furthermore, in α/β hydrolase all the catalytic triad residues are located in the loop regions, while the acid Asp^{75} in *E. coli* thioesterase/protease I is located in helix α3. Also, the NH protons of both Asp^{75} and His^{158} exchanged relatively slowly, suggesting that these two residues are well protected from solvent. In contrast, we were unable to observe the H^N resonance of Ser^{11}, suggesting that Ser^{11} is rather dynamic. It will be interesting to further investigate the catalytic implication of the dynamics of the Ser^{11} residue. The loop region between residues Gln^{150} and His^{158} is also interesting as the NH protons in this region exchanged rather slowly. Many $d_{NN}(i, i+2)$ and $d_{\alpha N}(i,i+2)$ NOEs were observed. Therefore, this region appears to form some sort of structure and is well protected. The catalytic His^{158} is located at the turn of this region.

Another interesting feature of the secondary structure is the location of the five blocks of consensus sequences of this subclass of enzymes, namely residues 5–15 (B1), 43–55 (B2), 67–77 (B3), 114–126 (B4), and 153–160 (B5) (Upton and Buckley, 1995). Three of these blocks (B1, B3, and B5) contain the three catalytic triad residues, while the other two blocks are mainly α-helices. Most of the NH protons of residues in these five blocks appear to exchange slowly; thus, these five blocks are likely to form the protected central catalytic core.

In conclusion, the secondary structure deduced from the present NMR study suggests that even though *E. coli* thioesterase/protease I shares little sequence identity with the common lipases, its structure may share some resemblance to that of the central core structure of the α/β hydrolase. However, substantial differences do exist and a more detailed structure must be obtained before a definitive comparison can be made.

Acknowledgements

This work was supported by the National Science Council of the Republic of China, NSC86-2314-B-001-031 (T.-h.H.) and NSC86-2311-B-001-017 (J.-F.S.) and the Academia Sinica.

References

Barnes Jr., E.M. and Wakil, S.J. (1968) *J. Biol. Chem.*, **243**, 2955–2962.

Barnes Jr., E.M., Swindell, A.C. and Wakil, S.J. (1970) *J. Biol. Chem.*, **245**, 3122–3128.

Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M. and Kay, L.E. (1990) *J. Magn. Reson.*, **87**, 620–627.

Bodenhausen, G. and Ruben, D.J. (1980) Chem. Phys. Lett., **69**, 185–189.

Borgström, B. and Brockman, H.L. (1984) *Lipases*, Elsevier, Amsterdam.

Brady, L., Brzozowski, A., Derewenda, Z.S., Dodson, G., Tolley, S., Turkenburg, J.P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thim, L. and Menge, U. (1990) *Nature*, **343**, 767–770.
Breddam, K. (1986) *Calsberg Res. Commun.*, **51**, 83–128.
Cambillau, C., Longhi, S., Nicolas, A. and Matinez, C. (1996) *Curr. Opin. Struct. Biol.*, **6**, 449–455.
Cho, H. and Cronan Jr., J.E. (1993) *J. Biol. Chem.*, **268**, 9238–9245.
Cho, H. and Cronan Jr., J.E. (1994) *J. Bacteriol.*, **176**, 1793–1795.
Cho, H. and Cronan Jr., J.E. (1995) *J. Biol. Chem.*, **270**, 4216–4219.
Clubb, R.T., Thanabal, V. and Wagner, G. (1992) *J. Magn. Reson.*, **97**, 213–217.
Cooper, Z. and Bussey, H. (1989) *Mol. Cell. Biol.*, **9**, 2706–2714.
Dalgarno, D.C., Levine, B.A. and Williams, R.J.P. (1983) *Biosci. Rep.*, **3**, 443–452.
Grzesiek, S. and Bax, A. (1992) *J. Magn. Reson.*, **96**, 432–440.
Grzesiek, S., Anglister, J. and Bax, A. (1993) *J. Magn. Reson.*, **B101**, 114–119.
Grzesiek, S. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 185–204.
Ichihara, S., Matsubara, Y., Kato, C., Akasaka, K. and Mizushima, S. (1993) *J. Bacteriol.*, **175**, 1032–1037.
Jaeger, K.E., Ransac, S., Dijkstra, B.W., Colson, C., van Heuvel, M. and Misset, O. (1994) *FEMS Microbiol. Rev.*, **15**, 29–63.
Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R. (1979) *J. Chem. Phys.*, **71**, 4546–4553.
Kass, L.R., Brock, D.J.H. and Bloch, K. (1967) *J. Biol. Chem.*, **242**, 4418–4431.
Lee, Y.L., Chen, J.C. and Shaw, J.F. (1997) *Biochem. Biophys. Res. Commun.*, **23**, 452–456.
Liao, D.-I. and Remington, S.J. (1990) *J. Biol. Chem.*, **265**, 6528–6531.
Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989a) *Biochemistry*, **28**, 6150–6156.
Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989b) *J. Magn. Reson.*, **85**, 393–399.
Neidig, K.-P., Geyer, M., Gorler, A., Antz, C., Saffrich, R., Beneike, W. and Kalbitzer, R. (1995) *J. Biomol. NMR*, **6**, 255–270.
Ollis, D.L., Chea, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H. and Goldman, A. (1992) *Protein Eng.*, **5**, 197–211.
Pacaud, M. and Uriel, J. (1971) *Eur. J. Biochem.*, **23**, 435–442.
Pacaud, M., Sibilli, L. and Le Bras, G. (1976) *J. Biochem.*, **69**, 141–151.
Pathak, D., Ngai, K.-L. and Ollis, D. (1988) *J. Mol. Biol.*, **204**, 435–445.
Pathak, D. and Ollis, D. (1990) *J. Mol. Biol.*, **214**, 497–525.
Piotto, M., Saudek, V. and Sklenář, V. (1992) *J. Biomol. NMR*, **2**, 661–665.
Richardson, J.S. (1981) *Adv. Protein Chem.*, **34**, 167–339.
Robertson, D.L., Hilton, S., Wong, K.R., Koepke, A. and Buckley, J.T. (1994) *J. Biol. Chem.*, **269**, 2146–2150.
Rubin, B. (1994) *Nat. Struct. Biol.*, **1**, 568–572.
Santaniello, E., Ferraboschi, P., Grisenti, P. and Manzocchi, A. (1992) *Chem. Rev.*, **92**, 1071–1140.
Shaka, A.J., Barker, P.B. and Freeman, R. (1985) *J. Magn. Reson.*, **64**, 547–552.
Spencer, A.K., Greenspan, A.D. and Cronan Jr., J.E. (1978) *J. Biol. Chem.*, **298**, 675–680.
Spera, S. and Bax, A. (1991) *J. Am. Chem. Soc.*, **113**, 5490–5492.
Su, M.S. (1997) Structure and function of *Escherichia coli* thioesterase/protease I, M.S. Thesis, National Oceanian University, Keelung, Taiwan.
Thomas, T., Cooper, A., Bussey, H. and Thomas, G.J. (1990) *J. Biol. Chem.*, **265**, 10821–10824.
Upton, C. and Buckley, J.T. (1995) *TIBS*, **20**, 178–179.
Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.
Wang, H. and Zuiderweg, E.R.P. (1995) *J. Biomol. NMR*, **5**, 207–211.
Winkler, F.K., D'Arcy, A. and Hunziker, W. (1990) *Nature*, **343**, 771–774.
Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.
Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.*, **239**, 363–392.
Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.
Wittekind, M. and Mueller, L. (1993) *J. Magn. Reson.*, **B101**, 201–205.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY.